# Table of Contents

## Acknowledgements

## 1 Introduction

1.1 Background biology

1.1.1 Deoxyribonucleic acid

1.1.2 DNA as the carrier of genetic information

1.2 Genomic data

1.2.1 Gene expression data

1.2.2 Copy number data

1.3 State of the art in Statistical Genomics

1.4 Thesis organisation

## 2 Fused lasso classifier

2.1 Introduction

2.2 Methods

2.2.1 Gradient Ascent

2.2.1.1 Regions, Constraints, Null Space and Projection Matrices

2.2.1.2 Calculating the penalized gradient

2.2.1.3 Finding the optimum and upgrading the coefficients

2.2.2 Newton-Raphson

2.2.3 Computational aspects and cross-validation

2.3 Results

2.3.1 Multiple Myeloma dataset

2.3.2 Bladder Cancer Data

2.3.3 Simulation study

2.4 Discussion

2.5 Conclusion
# TABLE OF CONTENTS

## 5 dNET

5.1 Introduction ........................................ 82
5.2 Methods ........................................... 83
   5.2.1 Motivation ..................................... 83
   5.2.2 Modelling the difference between the phenotypic groups .... 84
   5.2.3 Testing for a difference ......................... 85
   5.2.4 Estimation of parameters ......................... 86
   5.2.5 Permutation p-values ............................ 86
5.3 Results .............................................. 87
   5.3.1 Application on simulated datasets .................. 87
      5.3.1.1 Simulation setup ............................ 87
      5.3.1.2 Simulation study 1 ........................... 88
      5.3.1.3 Simulation study 2 ........................... 88
      5.3.1.4 Simulation study 3 ........................... 89
   5.4 Application on TCGA cancer datasets ................. 90
5.5 Discussion and Conclusion ............................ 92

## 6 Discussion ........................................... 95

6.1 Fused lasso for copy number based classification ........... 95
6.2 Integration analysis ................................... 95
6.3 Detecting differential associations ........................ 96
6.4 Association analysis between two high dimensional sets of variables .... 96
6.5 Differential network analysis based on gene sets ............ 97
6.6 Multiple groups for differential analysis .................... 97

## 7 Summary .............................................. 98

Samenvatting .............................................. 100

References ............................................... 103

Appendix A ............................................... 113

Appendix B ............................................... 115

B.1 Simulation study setup .................................. 115
   B.1.1 Generating copy number data ...................... 115
   B.1.2 Generating gene expression data .................. 115
   B.1.3 dSIM detects correct regions ...................... 116
   B.1.4 dSIM corrects for baseline effect ................. 116
   B.1.5 Sensitivity of dSIM towards changes in lambda values . 117
B.2 dSIM test statistic ...................................... 118
B.3 Table S1 ............................................... 119
## Appendix C

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1</td>
<td>Simulation setup</td>
<td>137</td>
</tr>
<tr>
<td>C.1.0.1</td>
<td>Generating copy number data</td>
<td>137</td>
</tr>
<tr>
<td>C.1.0.2</td>
<td>Generating gene expression data</td>
<td>137</td>
</tr>
<tr>
<td>C.1.0.3</td>
<td>Case 1: Widespread weak associations</td>
<td>138</td>
</tr>
<tr>
<td>C.1.0.4</td>
<td>Case 2: Localized strong associations</td>
<td>138</td>
</tr>
</tbody>
</table>